Effects of Bias in the Auto Insurance Industry

By Jai Raju, Oscar Casas

Introduction

Auto insurance is an agreement between policyholders and the insurance company that protects policyholders against financial loss in the event of an accident or theft. In exchange for policyholders' paying a premium, the insurance company agrees to pay their losses as outlined in the insurance policy. Basic personal auto insurance is mandated by most states and provides consumers with some financial protection in case of an accident. Some questions have risen from premium calculations themselves and how different consumers get different prices.

In the United States, impoverished groups tend to have higher insurance rates for the same coverage in comparison to the middle/upper class. The reason for this is because accidents and fatalities occur a lot more frequently in poverty stricken areas. This isn't because impoverished individuals cause more accidents but rather the fact that they usually drive older cars with less safety equipment such as airbags or automatic braking, the roads and infrastructure is worse and there are less trauma centers in the area.

Since individuals want the lowest premiums, wealthy individuals have gone for companies that segment people down and identify the lower risk metric associated with their wealth, this means the lower income groups will be segmented in the opposite direction and have higher premiums than their counterparts. While this segmentation is legal, low income individuals will continue to find themselves paying more for the misfortune they already face. In a country where history has forced poverty and race to be strongly correlated, this practice further increases the systemic racism affecting minorities today.

Car insurance depends upon answers to the following questions from a driver/policy owner:

- What type of car do they drive?
- How many miles a year do they drive?
- Where do they live?
- What is their driving history?
- Who will drive the car?
- Were there any gaps in car insurance coverage?

One can argue that the above data points can impact the potential risk of a claim that insurance is trying to cover. Therefore the premium can be mathematically designed as a function of the above variables.

$$prem = f(carType,\ milesDriven,\ location,\ numAtFaultAccidents,\ numOfMovingViolations,\ drivers, age)$$

All Auto Insurers indicate using the above variables in computing the premium in their Q&A.

Arguments

Variable 1: Credit Score as an indicator of auto risk

While the questions insurance companies ask, noted above, seem reasonable to use for premium computation, the formula is more complicated than it seems and actually takes a lot more data.

The insurers also get consumers credit score as a variable distilled from their social security number. Some insurers (StateFarm for example) make note of it in their Q&A and some (GEICO) dont. The use of credit scores in determining the auto risk behaviors of consumers and thereby the accidents does not seem reasonable as the correlation is not due to credit score itself but other variables such as people with low credit score tend to live in impoverished neighborhoods with bad infrastructure leading to more accidents.

Note: While the use of all the variables except Credit Score at the outset seems reasonable, note that there is no scientific proof for any of them to be used in premium computation.

The whole insurance industry uses observational data to summarize statistically the impact of these factors on a potential loss from a driver. While these variables are correlated to the loss there is no proof that these variables and their impacts are the actual cause of the loss.There is **no reasoning of the counterfactuals**.

Credit Score as an indicator of auto risk: Investigation

The purpose of Auto insurance is to protect consumers against a loss due to the operation of their vehicle, as such, it is protecting consumers from a mistake made while driving their vehicle. However, insurers not only use credit score but also weigh it as their most important variable in computing the loss. It can be argued that the Auto insurance companies correlate bad credit scores to claims and thereby to accidents. Here is why that line of thought is flawed:

a. Credit scores were created and are used for underwriting and pricing a variety of credit instruments and to measure the likelihood that an individual will repay their loans in full and on time. It is not an outcome of an experiment showing bad credit scores cause

accidents. Correlation does not mean causation and the auto insurers use of extraneous variables is potentially harming individuals and communities.

b. There is science and research that shows people of color disproportionately have lower credit scores. A 2012 study by the CFPB examining credit scores for about 200,000 consumers found that the median FICO score for consumers in majority minority zip codes was in the 34th percentile, while it was in the 52nd percentile for zip codes with minority minority populations as can be seen in figure 1. As a second example, consider figure 2 which is a chart generated by a credit processing company named Shift. The chart shows the average credit score for African American people as 644 compared to that of white people whose score is 734 even though the distribution of the population is not similar. Also the racial wealth gap didn't happen by accident. It was caused by centuries of slavery, redlining, and discrimination. For example, the practice of redlining was invented by the Federal Housing Administration, which refused to guarantee home loans made in African American communities, thus depriving them of the ability to accumulate wealth through homeownership. During the early years of Social Security, unemployment insurance, and the minimum wage, these programs did not cover domestic and agricultural workers — two of the most significant occupations for African-Americans thus perpetuating bias into the credit scoring system.

	Educational Median	FICO Median	Vantage Median	Difference	Educational Median	FICO Median	Vantage Median	Difference
	Majority Minority Areas				Low Minority Areas			
Educational vs. FICO	36	34	-	2	54	52	-	2
Educational vs. Vantage	35	-	35	0	53	-	53	0
FICO vs. Vantage	-	34	37	3	-	53	55	2

Figure 1:
Source: Consumer Financial Protection Bureau, Analysis of Differences Between Consumer- and Creditor-Purchased Credit Scores, at 18, Sept. 2012 [3]

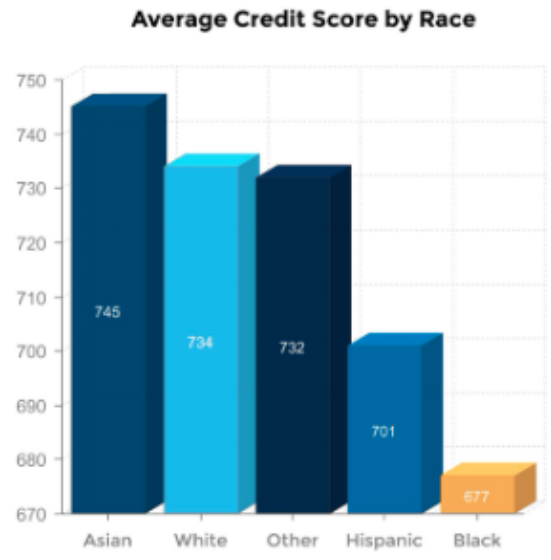

Figure 2: Average credit by race

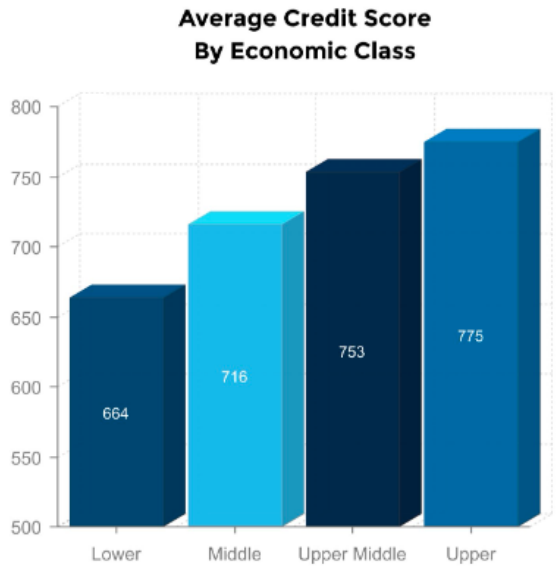

Figure 3: Average credit by Economic class

source : https://shiftprocessing.com/credit-score/

c. Credit score is determined by private, for-profit, publicly traded companies, whose goals are to turn a profit. Plus, Credit bureaus are under no legal requirement to be accurate, and the current credit reporting bureaus make a tremendous amount of mistakes at the consumers' cost. A 2013 Federal Trade Commission study of the U.S. credit reporting

industry discovered that 5% of consumers had errors. This disproportionately affects the poor as they cannot afford paying lawyers to get this corrected. A Congressional Research Service report stated that - consumers sometimes find it difficult to advocate for themselves when credit reporting issues arise because they are not aware of their rights and how to exercise them.

d. A federal report on the state of economics named Economic Well-Being of the U.S. Households in 2020 - May 2021 points that In 2020, 18% of Black Americans had no credit score, compared to 15% of Latinos, 13% of white Americans and 10% of Asian Americans[4]

Based on our investigation and the amount of evidence, we can conclude that there is a bias built into consumer scores and the fact that there is no evidence that credit score is related to bad driving through proxy, we think it is ethically and morally not correct to use credit score for scoring auto insurance risk.

Variable 2: Race subtly hidden in driving history

According to Pew Research Center, one of the areas in which race-based discrimination has a lengthy history is insurance where some pracitces, such as race-based premiums were common for over 250 years. Chlora Lindley-Myers, the Secretary-Treasurer of NAIC, notes that although the industry has eliminated many forms of direct racial discrimination, she points out that subtle, less obvious forms of discrimination remain[5]

Driving history is one such avenue where race based bias is hidden. In her book, Kelsey Shoub, an assistant professor of political science at the University of South Carolina and one of three co-authors of *Suspect Citizens: What 20 Million Traffic Stops Tell Us About Policing and Race* (Cambridge University Press, 2018) reports the following:

- Blacks were 63 percent more likely to be stopped even though, as a whole, they drove 16 percent less. Taking into account less time on the road, blacks were about 95 percent more likely to be stopped.
- Blacks were 115 percent more likely than whites to be searched in a traffic stop (5.05 percent for blacks, 2.35 percent for whites).
- Contraband was more likely to be found in searches of white drivers.

In another study, Pierson, E., Simoiu, C., Overgoor, J. et al. A large-scale analysis of racial disparities in police stops across the United States. Nat Hum Behav 4, 736–745 (2020)[6], the researchers show evidence that, relative to their share of the residential population, that black drivers were, on average, stopped more often than white drivers. The researchers use the veil-of-darkness test to show the evidence for bias in police stops and the actions subsequently.

Driver's driving record, which includes number of times the drivers are pulled over or the number of times a driver gets a citation are used as strong factors when determining premiums. The insurance companies draw statistical information based on these citations and generalize that the African-Americans are a lot more riskier drivers. There is no effort on the insurance companies' part to adjust for the disproportionate citations people of color receive. A Forbes advisor article, Jason Metz, Penny Gusner, *How Much Do Car Insurance Rates Go Up After A Speeding Ticket?*[7] notes that the national average car insurance rate increase for drivers who get a

speeding ticket is 24%, or nearly $380 more a year. And a rate increase due to a ticket can follow you around for many years.

Research :

So far, we have provided commentary based on the work done by other researchers. In this section we present our original research to show how violations affect the premiums. We went through many rate manuals for many insurance companies to find evidence on how rates vary by number of driving violations can increase premiums. We chose GEICO's rate manual[15] for the state of IL for our analysis. We used The System for Electronic Rates & Forms Filing (SERFF) to do our search of the rate filings which is a Department of Insurance requirement. We used the rate manuals we found at SERFF in the form of PDFs and carefully reconstructed the data into a spreadsheet in order to produce the charts and extract information out of it and is provided as a link[19]

For drivers of age 18, the Med Pay coverage has a premium surcharge of 15.5 % for one violation to 51.59% for 4 violations in the previous 12 months as shown in Figure 4. Similarly, for a 25 year old driver, the Med Pay coverage has a premium surcharge of 9.73% for one violation to 44.02% for 4 violations in the previous 12 months as shown in Figure 4. This trend continues for all ages and get worse for other coverages like Property Damage, which has a surcharge range of 21% - 68%. What this means is that a Black driver is 63% more likely to pay a 20-30% more premium due to the bias in policing and insurance companies do not do anything about it. We have provided the entire rate table for Violations from Geico as a link[16]

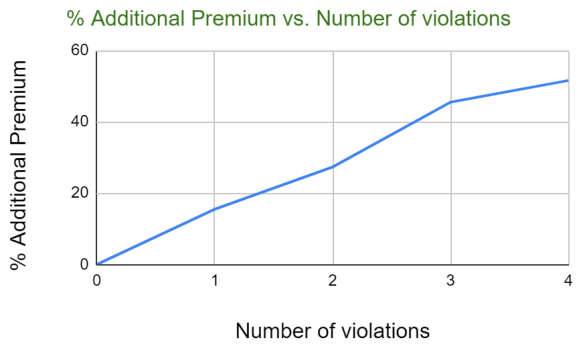

Figure 4: Med Pay Coverage for 18 year old drivers

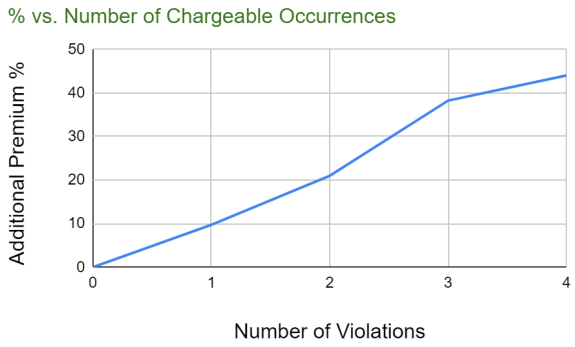

Figure 4: Med Pay Coverage for 25 year old drivers

Insurance companies use a term called “Tier”, which is used for protecting their business’ secret sauce, they don't disclose what goes in it. As one of the authors of this paper has insurance industry experience and has first hand experience in building the insurance models, we are able to say that the Tier variable is a combination of Credit score and other variables. We won't, therefore, be able to pull out the premium effects of credit scores. We provide a link to the aforementioned GIECO rate table which has the Tier rates.

Ethical Analysis

The Belmont report identifies three basic ethical principles that should guide researchers in making the best ethical judgements in human subject research. While there is no direct use of human subject in the context we discussed above, it certainly can be applied to the context on ethical grounds. We chose the Principle of Justice for this discussion.

The Justice principle advocates fair treatment for all and a fair distribution of the risks and benefits of the research. It forbids exploitation of vulnerable people (for instance, economically disadvantaged) or those who are easily manipulated as a result of their situation. This principle is violated in both the use cases discussed above.

In the first case where the credit score is used to calculate the driver's risk, the insurance companies are not treating the lower credit scores population the same as the higher credit score population. There is evidence that lower credit score population represents African-Americans at disproportionately high levels. Furthermore, the lower credit score population also generally comprises a population with lower financial resources. From the rate tables we see that lower credit scores result in higher premiums. That fact along with segments of population by race and economic class demonstrates unfair treatment and unfair distribution of risk and thus violate the Principle of Justice. It is clear that those who need the most help (African-Americans and the poor population) are not getting it. Further complicating the stride for a level the playing field and moreover not giving them a fair chance at life for no fault of theirs.

In the second case, where the police citations are used to compute the risk, the evidence clearly points at racial bias in police citations, which is one of the most influential variables in premium calculations (as seen in the rate tables). This indirect means of racial bias perpetuates unfair treatment of people of color and again violates the principle of Justice. From the perspective of Justice these calculations fail, as a country riddled with systemic racism further pushes in the wrong direction by giving misfortuned another hurdle along the way.

Present Day Regulation

In 1994, the fight against racial discrimination scored a big win when the National Association of Insurance Commisions (NAIC) Statistical Task Force developed statistical reporting guidelines. One of the guidelines' primary purposes was to investigate the use of zip code data of auto insurance companies and any identifying racial discrinimation that might have been employed with such data. Uncoincidentally big corporations such as the Fair Isaac Corporation and ChoicePoint introduced the first credit-based auto insurance policies. Since then the fight between racial descrimination and how data can be used for it has become murkier and getting targeted regulations has become ever more complicated. Historically this is evident from the results of NAIC Affordability and Availability Task Force 1998 final report. The report, *Improving Urban Insurance Markets: A Handbook on Available Options* included recommendations for legislators to consider, but more importantly held that racial descrimination in the auto insurance industry was most definitely prevalent. In some cases the report even found that identical policies were priced differently for African Americans than White individuals. The problem was that although the discrimination was present, the source of the descrimination was not evident to the researchers and moreover it didn't constitute *unfair* racial discrimnation laws. This conclusion, implying that there are such things as fair racial discrimnation.

This seems to be an ongoing issue in the US, where companies' data has an infinite number of ways to have racial bias but regulators struggle to tackle them due to their complexity, opaqueness and sometimes even the legislators motivation on the subject. Some regulators such as Texas Governor Rick Perry, signed amendments like the S.B. 14 which made race-based insurance policies illegal, but these laws did little to stop the discrimination since you don't need

someone race to racially discriminate. In 2007, the Federal Trade Commission study concluded "that credit-based insurance scores are effective predictors of risk under automobile insurance policies." Further, "Scores have only a small effect as a 'proxy' for membership in racial and ethnic groups in estimating insurance risk, remaining strong predictors of risk when controls for race, ethnicity and income are included in risk models." (Campbell, 8) This conclusion, if true, would be relatively sound, but in present day algorithms, credit is one of the top proxies for premiums and moreover the math ignores key statistical methods. The FTC claims seem to hint that because credit is a strong predictor when race, ethnicity and income are controlled that there is little evidence of racial bias. This assumes that just because a credit predictor held constant to race shows strong results that there is no discrimnation. The truth is that just because credit increases when race is held constant, does not mean that race stays constant as credit increases. As discussed, it is minorities that tend to have lower credit scores so regardless of the risk metric, it still ends in racial discrimination.

While the fight for racial equity among auto insurance policies looks gloomy, some strides have been taken. In 2019 Rep. Bonnie Watson Coleman introduced H.R. 3693, the Prohibit Auto Insurance Discrimination (PAID) Act, which aims to prohibit auto insurance using certain proxies to determine insurance rates and eligibility. These proxies consisting of :

- gender
- level of education
- occupation
- employment status
- home ownership status
- zip code or adjacent zip codes
- census tract
- marital status
- credit score or credit-based insurance score
- consumer report;
- previous insurer

- prior purchase of insurance of a consumer

Rep. Rashida Tlaib introduced another bill H.R. 1756, the Preventing Credit Score Discrimination in Auto Insurance Act. This Bill would prohibit the use of credit score, credit reports and other consumer information to determine premiums (Campbell, 9). While both these bills aim to solve the problem at hand, and some states such as California, Hawaii and Massachusetts already have some laws in place to protect consumers, the fight has a long road ahead to get to the route of the problem (Heller). With more data available, auto insurers can just revert back to their 1995 strategy and find another proxy for racial correlation to base their premiums off.

Lastly, while the fight for credit score and other proxies is on the front lines, police history remains unregulated and unchanged. Auto insurers base their premiums off police stops and history and in a society that targets minorities on the road, it's just another proxy to increase rates. No regulation has attempted to tackle this issue. The reason being that it seems common sense that people who get caught speeding or ticketed are at higher risk, the problem begins when minorities get stopped and ticketed at a disproportionate rate.

Privacy Policy Example

To understand the potential privacy infringements auto insurers inflict on consumers, a deep dive into one of the largest companies, GEICO, privacy policy yields worrying results. Geico starts its privacy policy by describing its data collection practices. It collects from the user:

- Name
- Address

- Phone number
- Email address
- Social Security number
- Driver's license number
- Date of birth

This seems like the necessary requirements for insurance. The only overreach that could be argued at this point is that insurance companies only need one identification number, so using just a drivers license should be sufficient. At this point as the user is the one inputting the data, the data is given consensually. However, auto insurance companies have made a practice of changing the data they collect and it could be an infringement on the interrogation clause under Solove's Privacy Taxonomy. For example, there has been historical proof of these companies using zip code data to discriminate, but users have to put that data in if they want to have the ability to drive and have mobility. This raises concerns about the possibility of auto insurance companies requesting more data with no limitations.

While most of the data may not lend itself to discrimination directly, the problem is it can be used to access more data. Geico states that it can collect third party data from consumer reporter agencies such as motor vehicle reports, claim history and credit information. This data isn't explicitly stated in the input for a quote and a reasonable user would not be aware of this. This lack of transparency is an issue when it comes to consent from the user. Lastly, Geico has its own cookie policy where it states it uses third party cookies. Geico would be able to know users' behaviors even when they are not on their website which goes against Solove's taxonomy on surveillance ("A Taxonomy of Privacy - ORG Wiki"). While Geico states its for marketing purposes, there is no explicit mention that the data collected can't be used to affect users premiums. Overall the data collected by Geico infringes on interrogation by giving users no

alternatives, surveillance by using users data outside of Geico's scope and has a lack of user consent as a reasonable user would not be aware of what Geico is collecting.

The problems don't stop at data collection however, the use of the data in its algorithm is where the real problems begin to manifest. Geico uses data aggregation to produce a risk metric for their quote. This aggregation imposes on Solove's Taxonomy of combining data to get a portrait of the person at hand ("A Taxonomy of Privacy - ORG Wiki"). By using all the data Geico has access to collect they can segment people confidently by race, income, gender and even more granular subsections. This can then be used against the user to increase their rates. While it is illegal to ask for race on the data collection phase, there are proxies such as credit, zip code and police stop history that can segment customers with high levels of confidence. These segments yield strong results for both auto insurance companies and for racial discrimination.

Geico's sharing of information is quite unrestricted and with these segments created can be quite damaging if given to the wrong hands. Geico states that it can share data with organizations performing a business for Geico or in the event of a business transaction. This means third parties can get access to Geico's segmentation and risk assessments which if racially discriminatory will be used downstream. This ability to sell or transfer the data so seamlessly breaks Soloves rule of secondary use and distortion where the data is being used for a purpose it wasn't meant to ("A Taxonomy of Privacy - ORG Wiki"). The data is also being generalized in a risk metric that may be detrimental to certain populations around the US. Geico's privacy policy has a lot of issues that primarily could be resolved by using less data to segment customers and limiting the sharing to its bare necessities. Until auto insurance companies stop trying to bypass

consumers' consent, privacy and right to not be defined by their segmentations, we will continue to see racial descrimination and impoverished groups negatively affected.

Positionality and Reflexivity Statements

1) **Oscar:**

Throughout the research and writing process of this assignment it really made me think about the impact of systematic racism on downstream processes of Americans' lives. We hear painfully naive arguments about how systematic racism no longer exists because the legislature and laws are equal for all diversity groups, and usually the arguments against this position are qualitative in nature but this is an extremely quantitative, concrete example of how minorities are at a huge disadvantage when it comes to social mobility and ability to save. Writing this document made me realize just how companies (either by choice or by negligence) have the ability to target underprivileged groups indirectly. You don't need their race to segment down to it. In a country where minorities have been shoved into lower income, high crime neighborhoods their access to capital, education and affordable insurance has been impeded by companies using the segmentations that have racial biases. Lastly, I also realized that while as data scientists we pursue high granularity, we need to be mindful of when the granularity is impacted by unjust forces. It's easy to find correlation, but sometimes that isn't enough and we should be ready to keep these questions in the back of our minds with whatever project we may encounter.

2) **Jai:**

As one of the authors of this paper, I'd like to disclose my positionality as my worldview has offered unique opportunities to ask questions pertaining to the use of credit score and citations in the risk models.

I am an Asian-Indian, I lived a good part of my life in extreme poverty and out of poverty. I have been an executive in a large insurance company with responsibilities for a Data Science department.

I was in an accident where a white woman collided with my car sideways. The officer who investigated only took the account of the other driver who was a white woman and gave me a citation for something that was not my fault and could be proved. I was not even given a chance to speak, let alone explain what happened. I thought I was discriminated against (perhaps, due to my color or the foreignness of my appearance).

I recognize these and many other factors shape my understanding of discrimination due to color, gender, poverty, socio-economic classes and the interpretation of research data.

References

1. Nelson, Ashlyn Aiko. “Credit Scores, Race, and Residential Sorting.” *Journal of Policy Analysis and Management*, vol. 29, no. 1, 2010, pp. 39–68. *JSTOR*, http://www.jstor.org/stable/20685167. Accessed 26 Jul. 2022.
2. “Past Imperfect: How Credit Scores and Other Analytics ‘Bake In’ and Perpetuate Past Discrimination.” *Nclc.Org*, National Consumer Law Center, 2016, www.nclc.org/images/pdf/credit_discrimination/Past_Imperfect050616.pdf.
3. “Analysis of Differences between Consumer- and Creditor-Purchased Credit Scores.” *Consumerfinance.Gov*, Consumer Finance Protection Bureau, Sept. 2012, files.consumerfinance.gov/f/201209_Analysis_Differences_Consumer_Credit.pdf.
4. “The Fed - Economic Well-Being of U.S. Households in 2020 - May 2021 - Banking and Credit.” *Board of Governors of the Federal Reserve System*, Federal Reserve, May 2021, www.federalreserve.gov/publications/2021-economic-well-being-of-us-households-in-2020-banking-and-credit.htm.
5. NAIC. "Race & Insurance | NAIC Summer National Meeting 2020." YouTube 20:28.
6. Pierson, E., Simoiu, C., Overgoor, J. *et al.* A large-scale analysis of racial disparities in police stops across the United States. *Nat Hum Behav* 4, 736–745 (2020). https://doi.org/10.1038/s41562-020-0858-1
7. Forbes advisor article, Jason Metz, Penny Gusner, How Much Do Car Insurance Rates Go Up After A Speeding Ticket? https://www.forbes.com/advisor/car-insurance/speeding-ticket-rate-increase/

8. Heller, Doug. “Systemic Racism in Auto Insurance Exists and Must Be Addressed By Insurance Commissioners and Lawmakers.” *Consumerfed.Org*, CFA, 17 June 2020, consumerfed.org/press_release/systemic-racism-in-auto-insurance-exists-and-must-be-addressed-by-insurance-commissioners-and-lawmakers.
9. Powell, Lawrence. “Big Data and Regulation in the Insurance Industry.” *SSRN Electronic Journal*, 2017. *Crossref*, https://doi.org/10.2139/ssrn.2951306.
10. “Text - H.R.3693 - 116th Congress (2019–2020): Prohibit Auto Insurance Discrimination Act.” *Congress.Gov | Library of Congress*, www.congress.gov/bill/116th-congress/house-bill/3693/text?q=%7B%22search%22%3A%5B%22consumer%22%5D%7D&r=16&s=1. Accessed 28 July 2022.
11. Campbell, Eryn, et al. “Milestones in Racial Discrimination within the Insurance Sector.” *NAIC.Org*, NAIC, 1 Aug. 2020, content.naic.org/sites/default/files/cipr-report-milestones-racial-discrimination.pdf.
12. “Privacy Request and Policies | GEICO.” *Geico*, www.geico.com/privacy. Accessed 28 July 2022.
13. “A Taxonomy of Privacy - ORG Wiki.” *Open Rights Group*, wiki.openrightsgroup.org/wiki/A_Taxonomy_of_Privacy. Accessed 28 July 2022.
14. Nelson, Ashlyn Aiko. “Credit Scores, Race, and Residential Sorting.” *Journal of Policy Analysis and Management*, vol. 29, no. 1, 2010, pp. 39–68. *JSTOR*, http://www.jstor.org/stable/20685167. Accessed 26 Jul. 2022.
15. GC Manual Part 2.pdf
16. Geico - Rate table for Minor and Speeding Violations - 1429